Original Research

DOI: doi.org/10.47363/AMR/2015(3)129

Psychometric properties of the SF-12, penn shoulder score, and visual analog scale-pain for rotator cuff retears

Jyothi Menon[1], Joseph C. Cappelleri[2], Jack Mardekian[2], Nicholas J. Vendetti[3], Marko Mychaskiw[4], Joseph Thomas III[5]

[1]Purdue University, College of Pharmacy and Regenstrief Center for Healthcare Engineering, Center for Health Outcomes Research and Policy
[2]Istatistics Pfizer, Inc
[3]Global Health and Value, Pfizer Inc
[4]Disease Area Lead Specialty Care Medicines Development Group, Pfizer Inc.
[5]Purdue University and College of Pharmacy and Director Regenstrief Center for Healthcare Engineering, Center for Health Outcomes Research and Policy

Address for correspondence:
Jyothi Menon,
Purdue University, College of Pharmacy and Regenstrief Center for Healthcare Engineering, Center for Health Outcomes Research and Policy
jyothimenon05@gmail.com

Received: May 26, 2015
Accepted: June 23, 2015
Published: July 07, 2015

ABSTRACT

Purpose: We assessed reliability, responsiveness, and sensitivity of the Short Form-12 (SF-12), Visual Analog Scale (VAS) Pain score, and Penn Shoulder Score (PSS) in patients with surgical repair of rotator cuff tear. **Methods:** The SF-12, PSS, and VAS-Pain were administered at a screening visit and at six follow-up visits over 1 year. Internal consistency, test-retest reliability, standardized effect sizes, and corrected item-total correlations were calculated and cumulative distribution function plots were constructed. **Results:** Cronbach's alpha coefficients exceeded 0.70 for PSS Pain and Function subscales but not for PCS-12 or MCS-12 of the SF-12. The PSS Function, PCS-12, MCS-12, and VAS-Pain gave intraclass correlation coefficients (ICCs) between 0.60 and 0.80, while PSS Pain and Satisfaction subscales gave ICCs between 0.44 and 0.73. Standardized effect sizes between week 52 and baseline were between 1.5 and 3.5 for all scales, apart from the MCS-12 effect size of 0.005. Effect sizes between retear and no retear groups were between 0.3 and 0.7 for all scales, apart from MCS-12 effect size of 0.10. Cumulative distribution functions between retear and no retear groups revealed visual separation between changes in the PSS total score, PCS-12, and VAS Pain score, but no separation in the MCS-12. **Conclusions:** Cronbach's alpha for PCS-12 and MCS-12 scales were below the generally accepted level of 0.70. The PCS-12, MCS-12, PSS total, PSS Function subscale, and VAS-Pain showed moderate test-retest reliability while PSS Pain and Satisfaction subscales showed fair test-retest reliability. All scales except the MCS-12 showed excellent responsiveness and discernible sensitivity.

KEY WORDS: Reliability, Responsiveness, SF-12, PSS, VAS-Pain, Cumulative Distribution Function

INTRODUCTION

Background

The rotator cuff is a network of four muscles and their tendons that form a covering around the head of the humerus or upper arm bone [1]. When one or more of the rotator cuff tendons is torn, the tendon no longer fully attaches to the head of the humerus. Rotator cuff tears are associated with muscle overuse, natural aging processes, tobacco use, and diabetes mellitus [2, 3]. Rotator cuff tears are more common in patients involved in physical work and sports [4]. When non-surgical options do not decrease pain or improve function, surgery is recommended [5].

There are three options for rotator cuff tears surgery: open surgery, mini-open surgery, and all-arthroscopic surgery [6]. Arthroscopic surgery has proven successful in treating rotator cuff tears [3] and is preferred over open surgery because arthroscopic surgery is associated with fewer postoperative complications, including pain and scarring, than open surgery [3, 7]. A common complication after rotator cuff repair surgery is rotator cuff retears [8]. Size of the tear, age, improper rehabilitation, degree of muscular atrophy, and degree of fatty infiltration of the cuff muscle have been identified as risk factors for retears [2, 8, 9].

Surgery on rotator cuff tears is intended to improve daily activities of patients, and improve their health-related quality of life [10]. Effects of arthroscopic surgeries on health-related quality of life of patients have been examined using self-reported, patient-based instruments [10-12]. Generic health-related quality of life instruments are useful in determining health status of any individual and condition-specific instruments ascertain disability or Function due to a particular condition. Shoulder functional outcomes have been examined using Penn Shoulder Score (PSS) [13, 14] and using generic instruments including the SF-12 and SF-36 [10, 15], and the VAS-Pain [16, 17].

We found no reports that compared the SF-12, PSS, and VAS-Pain among a population suffering from rotator cuff tears. Since there is no established gold standard for shoulder health-related quality of life outcomes, comparison of these instruments' reliability, validity, and responsiveness in individuals suffering from rotator cuff tears would be valuable. This study assessed internal reliability, test-retest reliability, responsiveness, and sensitivity for the SF-12, PSS, and VAS Pain score for subjects who underwent rotator cuff tear surgery.

METHODS

Data Sources

Data for analysis was obtained from a clinical trial (Protocol number B1861001) conducted by Pfizer, Inc. with the study number of 3202V1-1000 and with NLM identifier: NCT00739947 from clinicaltrials.gov [18]. Iannotti and colleagues used the data from this trial to evaluate time to failure after rotator cuff repair and concluded that retears primarily occurred between six and twenty-six weeks after rotator cuff repair [19]. Data for the current study included information on the race, sex, and age of individuals, whether an individual suffered a retear during the post-operative period, and responses on the SF-12, PSS, and the VAS-Pain scale. Cases with missing data on any of the study variables were excluded from the analysis.

Sample Selection

Sample inclusion criteria were ages between 21 years and 75 years inclusive with a full-thickness rotator cuff tear between one and four centimeters, as estimated on magnetic resonance imaging (MRI) taken within 3 months before surgical repair. Subjects were required to have a normal active range of motion in their contralateral shoulder.

Sample exclusion criteria included previous, shoulder arthroscopy, acromioplasty, previous rotator cuff repair, fracture of the shoulder joint; tears of the subscapularis or labral pathology; moderate or severe degenerative glenohumeral arthritis, avascular necrosis or chondrocalcinosis, or, rheumatoid arthritis affecting the shoulder joints; Stage 3 or 4 fatty infiltration (according to the Goutallier 7 grading scale) of their rotator cuff muscles on MRI taken within 3 months of surgical repair; shoulder instability in either shoulder; or treatment with corticosteroid injections in the shoulder under study within 3 months before surgical intervention.

Individuals who suffered either a local or systemic infection that would preclude arthroscopy or those who were either unwilling or unable to undergo examination with closed MRI (e.g., because of claustrophobia or the presence of a pacemaker or an automatic cardioverter defibrillator) were excluded. Subjects unable to complete functional evaluations prior to surgical repair were also excluded from analysis.

Data Collection

Subjects scheduled to undergo arthroscopic surgical repair for their rotator cuff tears provided informed consent and were screened within three months before surgical repair. The SF-12, PSS and VAS-Pain were administered to subjects at the screening visit and at each post-operative visit beginning at 6 weeks, 12 weeks, 16 weeks, 26 weeks, 39 weeks and 52 weeks after surgery.

Ethical Consideration

The clinical trial from which data for this study was derived, was in accordance with applicable laws and regulations including, but not limited to, the International Conference on Harmonisation (ICH) Guideline for Good Clinical Practice (GCP) and the ethical principles that have their origins in the Declaration of Helsinki Institutional Review Board. Approval for the current study was obtained from the Institutional Review Board at Purdue University, West Lafayette, Indiana.

Study Variables

Demographic Variables

Demographic variables including patient age, gender, and race were collected. Age was collected as a categorical variable with individuals younger than forty years grouped together, ages forty to fifty years grouped together, ages fifty-one to sixty years grouped together, and individuals older than sixty years grouped together.

Retear Variable

Individuals were examined for muscle retears at post-operative visits from visit three to visit nine using MRI assessments. Any subject who had at least one retear during this period was categorized as having a retear. A variable indicating retear was coded 'Yes' for a retear or 'No' for no retear.

Patient-reported Outcomes

The Short Form-12 Questionnaire (SF-12)

The SF-12 consists of 12 questions concerning physical and mental function and well-being. The domains include physical functioning, role physical, general health, mental health, social functioning, role emotional and vitality. There are two component summary scores - the physical component summary (PCS) score and the mental component summary (MCS) score that are generated after assigning individual weights to each domain, using scoring instructions provided by Ware, Kosinski and Keller [20].

Penn Shoulder Score (PSS)

The Penn Shoulder Score consists of three subscales: Pain, Satisfaction, and Function. The Pain subscale consists

of three items that address pain at rest, pain with normal activities, and pain with strenuous activities. A patient is awarded a maximum of thirty points for complete absence of pain. If a patient is not able to use the arm for normal or strenuous activities, zero points are awarded for that item. The Satisfaction subscale is assessed with a ten-point numeric rating scale ranging from "not satisfied" with a value of '0' to "very satisfied" with a value of "10". The Function subscale is based on a sum of twenty items, each with a four-point Likert scale. A patient is awarded sixty points if he or she can perform all activities without difficulty. Scores for each subscale were calculated using instructions provided by Leggin and colleagues [13].

Visual Analog Scale (VAS)-Pain

The VAS for Pain is a standard 100-mm visual analog scale that was used to obtain subject's self-assessment of Pain in the injured shoulder. A score of 0 indicates 'no Pain', while a score of 100 indicates 'worst possible Pain' [21].

Study Analyses

Data management and analysis was carried out using SAS for UNIX version 9.3 [22]. An a priori alpha level of 0.05 was used for all analyses. Cumulative distribution function plots for the scales were created using Stata for UNIX version 12.1 [23].

Internal Consistency

Cronbach's alpha for the SF-12, PSS, and VAS-Pain at baseline pre-operative visits and final post-operative visit were calculated. Nunnally's criterion of a Cronbach's alpha of 0.70 or higher was considered as a threshold for internal consistency reliability [24].

Test-retest Reliability

To assess test-retest reliability of the SF-12, PSS, and VAS-Pain, intraclass correlation coefficients were calculated for scale scores at weeks 39 and 52 of the post-operative visits. This time period was chosen as it was considered a relatively stable period with little to no change was expected in the health status of the patient. Intraclass coefficient correlations were determined using the Intracc macro for SAS developed by Hamer to determine test-retest reliability [25]. The macro calculates intraclass correlation (2,1) as discussed in Shrout and Fleiss [26]. Intraclass correlations (2,1) from a two-way random effects model where patients and time have random effects were examined, and the absolute agreement definition of concordance was computed [26, 27].

Item-level Validity

For multi-item scales, corrected item-to-total correlation (i.e., the item in question is removed from the total score) was performed for each item to determine if an item was part of the same construct (concept) as measured by the other items in the same scale.

Responsiveness and Sensitivity

Distribution-based methods allow for standardization among scales with different ranges and different scoring. One such distribution-based method includes estimating effect sizes [28].

For each patient-reported outcome, standardized effect sizes for responsiveness (within-subject change) were calculated by subtracting each individual's baseline score at pre-operative visit from the final score post-operatively at week 52 to calculate each change score, and then this mean change in score was divided by the standard deviation of the baseline scores to compute an effect size for responsiveness. Standardized effect sizes for sensitivity between retear and no retear groups were calculated by estimating the difference in mean changes from the baseline visit to the week 52 visit between the no retear group and the retear group, and then dividing this difference by the pooled baseline standard deviation.

Interpretation

Cumulative distribution function plots provide a graphical interpretation of responsiveness within groups and sensitivity between groups [28, 29]. Change from baseline of the patient-reported outcome scores is presented on the horizontal axis and the cumulative percent of patients experiencing up to that change is presented on the vertical axis.

For PCS-12, MCS-12 and PSS plots, changes between week 52 and baseline (change = week 52 score minus baseline score) were presented on the horizontal axis and the cumulative percent of patients who experienced that change or less were presented on the vertical axis. Positive changes in difference in mean scores between groups for PCS-12, MCS-12, and PSS represented better health.

For VAS-Pain, higher scores are indicative of greater pain. Individuals are expected to have lower scores (lesser pain) on the VAS-Pain after surgery as compared to prior surgery. To accommodate for this in VAS-Pain plot, changes between baseline and week 52 (change = baseline score – week 52 score) were presented on the horizontal axis and cumulative percent of patients experience that change in score or less was presented on the vertical axis. In the case of VAS-Pain, as was in the case of the PCS-12, MCS-12 and PSS, positive changes in the difference in mean scores between groups are favorable. Differences in the cumulative distribution functions between retear and no retear groups were tested with the Kolmogorov-Smirnov test [30].

RESULTS

Sample

A total of 199 subjects from 13 investigation sites were randomized in the clinical trial. Thirty-seven of the 199 individuals (18.6%) were not evaluable for various reasons (including withdrawal before 6 months of follow-up and violation of one or more inclusion or exclusion criteria), leaving a total of 162 individuals. Nine individuals did not have any demographic information leaving a sample of 153 individuals who had demographic information. A total of 22 individuals suffered retear and the remaining 131 individuals did not. Sample demographic characteristics are summarized in Table 1. Seventy-one individuals (46.41%) were older than 60 years of age, and 91 individuals (59%) were males, and 131 individuals (86%) were Caucasian.

Internal consistency reliability

Table 2 provides the Cronbach alpha coefficients for each of the scales based on available scores obtained at pre-operative baseline visit and final post-operative visit at week 52 for the total sample, individuals with retears, and individuals with no retears. Cronbach's alpha coefficient for the PCS-12 was 0.68 at baseline and at the week 52 post-operative visit for the total sample while Cronbach's alpha coefficient for the MCS-12 was 0.63 at baseline and 0.67 at week 52 for the total sample.

Table 1. Sample Demographic Distribution for total sample, individuals with no retears and individuals with retears

	Total (N=153)[a]		No Retear group (N=131)		Retear group (N=22)		
Demographic	Frequency	Percent	Frequency	Percent	Frequency	Percent	P-value[b]
Age							
Younger than 40 years	4	2.61	4	3.05	0	0.00	
40 to 50 years	24	15.69	21	16.03	3	13.64	0.482
51 to 60 years	54	35.29	46	35.11	8	36.36	
Greater than 60 years	71	46.41	60	45.80	11	50.00	
Gender							
Male	91	59.47	83	63.36	8	36.36	0.049
Female	62	40.52	48	36.64	14	63.64	
Race							
Asian	2	1.30	2	1.53	0	0.00	
Black African American	14	9.15	13	9.92	1	4.55	0.030
Caucasian	131	85.62	111	84.73	20	90.91	
Hispanic	6	3.92	5	3.82	1	4.55	

[a]Total number of individuals in the study were 162, with 9 individuals having missing data on the demographics
[b]From the chi-square test based on the null hypothesis of equal proportions

Table 2.Cronbach Coefficients for PCS, MCS, and Penn Shoulder Scores with Pain and Function subscales scores for baseline pre-operative and final post-operative visits for total sample, individuals with no retears and individuals with retear

	Total group (N=160)		No Retear group (N=139)		Retear group (N=21)	
Scores	Baseline pre-operative visit	5 2-week post-operative visit	Baseline pre-operative visit	52-week post-operative visit	Baseline pre-operative visit	52-week post-operative visit
SF-12	-	-	-	-	-	-
PCS-12	0.68	0.68	0.68	0.68	0.70	0.71
MCS-12	0.63	0.67	0.69	0.69	0.69	0.73
Penn Shoulder						
Score	0.81	0.84	0.81	0.85	0.82	0.81
Pain	0.85	0.86	0.84	0.86	0.86	0.87
Function	0.76	0.76	0.76	0.76	0.76	0.94

Note: Satisfaction subscale in Penn Shoulder Score and VAS-Pain scale are one-item scales; therefore, Cronbach's alpha cannot be estimated for them. Two individuals from a sample of 162 individuals with missing data on the scales were not included in the analyses.

Cronbach's alpha coefficients for the total PSS values for the total sample were 0.81 at baseline and 0.84 at week 52. The PSS Pain subscale had values of 0.85 and 0.86 at baseline and at week 52, respectively, and the PSS Function subscale had a value of 0.76 at baseline and at week 52 visits. Cronbach's alpha could not be estimated for PSS Satisfaction subscale or VAS-Pain as both of them have one item each.

SF-12 items that showed a sizeable correlation with MCS-12, including those on mental health, role emotional, general health, and social functioning, had high item-to-total correlation of at least 0.4. Similarly, SF-12 items that showed a sizeable correlation with PCS-12 including those on physical health, role physical, bodily pain, and vitality also had item-to-total correlation of at least 0.4. Pain, Satisfaction, and Function subscales of the PSS correlated meaningfully with the total PSS score with values higher than 0.4. Items within Pain and Function subscales also showed item-to-total correlation of greater than 0.4.

Test-retest Reliability

Table 3 provides the estimated ICCs (for the absolute agreement of concordance) between weeks 39 and 52 after surgery for component summary scores of the SF-12; total score, Pain, Satisfaction and Function subscales of the PSS; and the VAS-Pain scales. For PCS-12, the ICC was 0.62 for MCS-12, the ICC was 0.69. The ICC for the total PSS score was 0.69, 0.55 for the PSS Pain subscale, 0.48 for the PSS Satisfaction subscale, 0.73 for the PSS Function subscale, and 0.62 for the VAS-Pain scale.

Effect Sizes

Tables 4 and 5 provide baseline pre-operative scores and scores at final post-operative visit at 52 weeks, mean change in scores, baseline standard deviation, and (standardized) effect sizes for the component summary scores of the SF-12, total PSS scores and the PSS subscale scores, and the VAS-Pain scale scores. For the total sample, the effect size for the PCS-12 was 1.51, while the effect size for MCS-12 was 0.005; the effect sizes were 2.71 for the total PSS, 2.91 for the PSS Pain subscale, 3.35 for PSS Satisfaction subscale, 2.25 for the PSS Function subscale and 2.02 for the VAS-Pain scale (Table 4).

Table 3 Intraclass correlation coefficients (for absolute agreement) between weeks 39 and 52 to assess test-retest reliability for PCS- 12 and MCS-12, Penn Shoulder Scores (Pain, Satisfaction and Function subscales) and VAS-Pain scale

	Total (N=126)	No Retear group (N=110)	Retear group (N=16)
Scores	Absolute agreement	Absolute agreement	Absolute agreement
SF-12	-	-	-
PCS-12	0.62	0.62	0.63
MCS-12	0.69	0.79	0.48
Penn Shoulder Score	0.69	0.69	0.71
Pain	0.55	0.55	0.54
Satisfaction	0.48	0.44	0.73
Function	0.73	0.72	0.79
VAS-Pain	0.62	0.66	0.44

Table 4. Effect sizes for PCS-12 and MCS-12, Penn Shoulder Scores (total Pain. Satisfaction and Function subscales). and VAS-Pain scale at baseline pre-operative visit and after 52 weeks for total sample (N=127)

Scores	Baseline pre-operative mean scores	52-week post-operative mean scores	Mean change scores	Baseline Standard deviation	Effect Size
SF-12	-	.	-	-	-
PCS-12	37.96	51.09	13.13	8.68	1.51
MCS-12	54.57	54.63	0.06	10.19	0.005
Penn Shoulder Score- total	46.72	91.83	45.10	16.65	2.71
Pain	14.94	27.78	12.84	5.86	2.91
Satisfaction	1.70	8.89	7.19	2.15	3.35
Function	30.11	55.12	25.01	11.09	2.25
VAS-Pain[a]	56.15	5.88	-50.27	24.89	2.02

[a] Mean change was defined as Week 52 minus preoperative baseline level for SF-12 and Penn Shoulder Score. For VAS-Pain, mean change was defined as baseline level minus Week 52, Therefore, for all three sets of outcomes, more positive changes and effect sizes are more favorable.

Effect size between retear and no retear groups were also estimated for each scale (Table 5). The effect sizes between retear and no retear groups were estimated at 0.30 for PCS-12, 0.11 for MCS-12, 0.54 for total PSS, 0.45 for PSS Pain 0.71 for PSS Satisfaction , 0.42 for PSS Function and, finally, 0.51 for VAS-Pain.

Cumulative Distribution Function Plots

Figure 1 shows cumulative distribution function plots for the change scores on the PCS-12, MCS-12, PSS, and VAS-Pain. For the graph depicting change in PCS scores versus the percentage of patients experiencing change in the score, there was appreciable separation between retear and no retear groups for most change scores but not for the higher change scores (Kolmogorov-Smirnov test p-value=0.268). Approximately 60% of subjects in the retear group had scores of 10 points or less, versus approximately 30% of subjects in the no retear group. The cumulative distribution function plot for the MCS clearly had substantial overlap throughout between the retear and no retear groups for positive change scores (Kolmogorov-Smirnov test p-value = 0.914).

For the total PSS, despite the lack of statistical significance (p=0.159), there was evidence of clear separation between the two curves for the majority of values. Approximately 55% of the subjects had scores of 40 points or less in the tear group, relative to approximately 30% of subjects in the no retear group. The same type of separation can be noted for the VAS-Pain scale (Kolmogorov-Smirnov test p-value=0.18). Approximately 40% of subjects in the retear group had scores of 20 points or less, compared with 10% in the no retear group

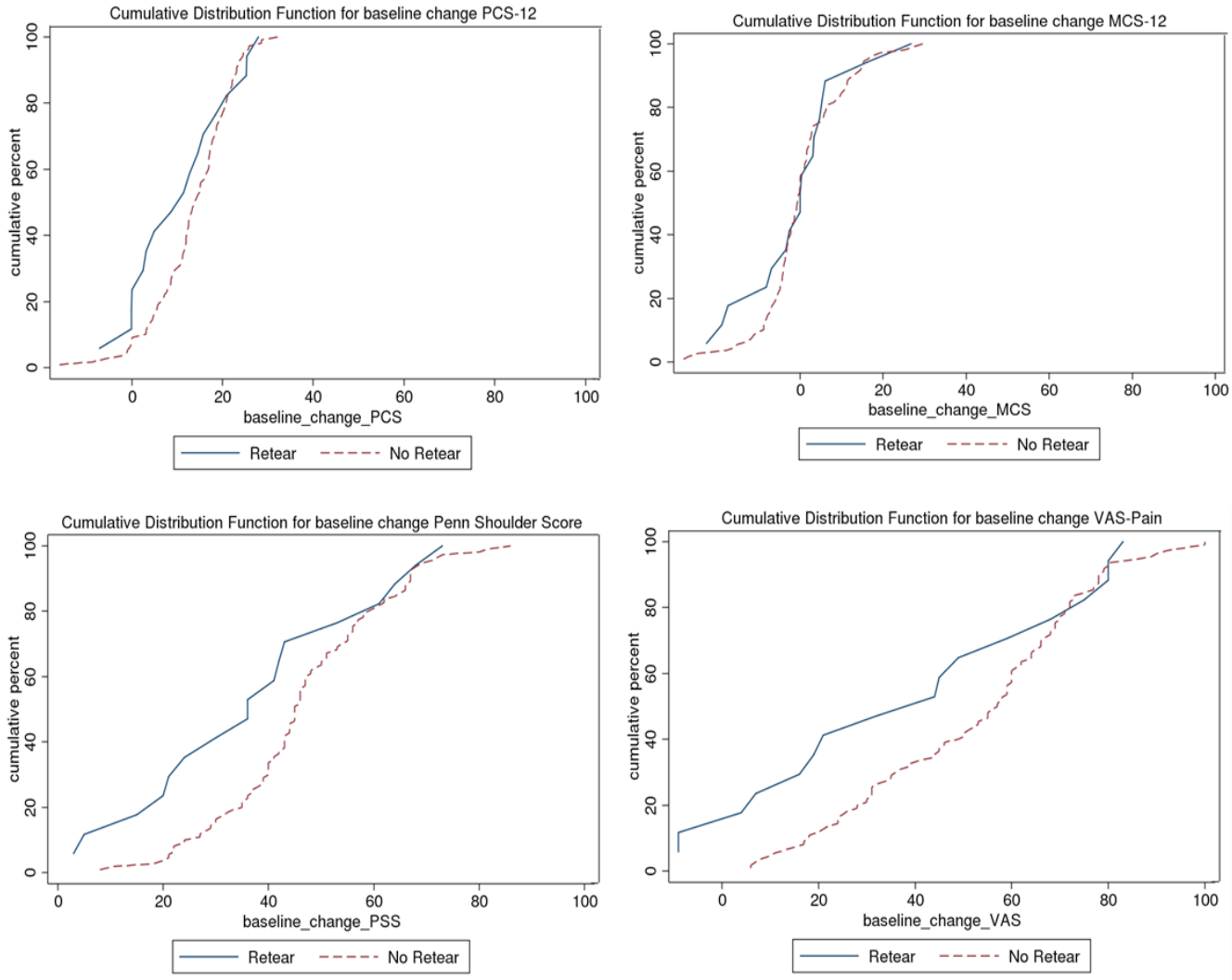

Fig. 1 Cumulative distribution function plots for PCS-12, MCS-12, Penn Shoulder Score, and VAS-Pain with two curves indicating individuals with retears and individuals without retears. In the horizontal axis a higher change in mean change scores for PCS-12, MCS-12, PSS, and VAS-Pain represent more favorable responses. The vertical axis is cumulative percentage of individuals experiencing that level of change or less in the respective patient-reported outcome scores.

Table 5. Effect sizes for PCS-12 and MCS-12, Penn Shoulder Scores (total and also Pain. Satisfaction and Function subscales), and VAS-Pain scale at baseline pre-operative visit and after 52 weeks for individuals with no retear injuries and with retear injuries

	No Retear group (N=110)					Retear group (N=17)					
Score	Baseline mean scores	52-week mean scores	Mean[a] chanee scores	Baseline standard deviation	Effect[b] Size of no retear group	Baseline mean scores	52-week mean scores	Mean[a] change scores	Baseline standard deviation	Effect[b] size of retear group	Effect size[c] between no retear and retear groups
SF-12	–	–	–	–	–	–	–	–	–	–	–
PCS-12	37.99	51.47	13.48	8.37	1.61	37.79	48.63	10.85	10.64	1.02	0.30
MCS-12	55.20	55.39	0.19	9.5	0.02	50.55	49.72	-0.84	10.50	0.08	0.11
Penn Shoulder											
scores	46.52	92.81	46.30	15.96	2.90	48.06	85.41	37.35	21.22	1.76	0.54
Pain	14.92	28.12	13.20	5.71	2.31	15.06	25.58	10.53	6.97	1.51	0.45
Satisfaction	1.63	9.03	7.40	2.11	3.51	2.18	8.06	5.88	2.40	2.45	0.71
Function	29.99	55.63	25.64	10.73	2.39	30.83	51.76	20.93	13.50	1.55	0.42
VAS-Pain	56.44	4.44	-52.00	24.53	2.12	54.23	15.17	-39.06	31.00	1.26	0.51

[a] Mean change was defined as Week 52 minus preoperative baseline level for SF-12 and Penn Shoulder Score. For VAS-Pain, mean change was defined as baseline level minus Week 52, Therefore, for all three sets of outcomes, more positive changes and effect sizes are more favorable.
[b] Effect size within group (responsiveness) is the difference in its mean scores between baseline and Week 52, divided by its baseline standard deviation.
[c]Effect size between groups (sensitivity) is the difference in their mean changes, divided by their pooled baseline standard deviation.

DISCUSSION AND CONCLUSIONS

To our knowledge, no study has reported reliability, responsiveness, and sensitivity of the SF-12 in a population that underwent arthroscopic surgery on rotator cuff tears. Cronbach's alpha for PSS total and its Pain, and Function subscales were higher than 0.70 in our study, indicating good reliability according to Nunnally's criterion [24]. However, PCS-12 and MCS-12 were not higher than 0.70 for the total sample. Estimates of ICCs have been given the following classifications: with less than 0 as poor; between 0 and 0.2 as slight agreement; between 0.21and 0.40 as fair agreement; between 0.41 and 0.60 as moderate; between 0.61 and 0.80 as substantial; and between 0.81 and 1.00 as almost perfect [31]. The test-retest reliability coefficients in the current study were 0.62 for the PCS-12 scale and 0.69 for the MCS-12 scale, indicating that they had "substantial" reliability. Ideally, we would have liked to see ICC values above 0.70. The test-retest reliability coefficients from a general population obtained by Ware and colleagues while developing the SF-12 instrument was higher with 0.89 for PCS-scale and 0.75 for MCS-scale [20].

For the Penn Shoulder Scale, Leggin and colleagues reported a test-retest reliability of 0.88 for the Pain subscale and 0.93 for Satisfaction subscale as well as for Function subscale for a cohort of individuals who suffered from various shoulder pathologies and were undergoing physical therapy [13]. In the current study, test-retest reliability estimates were estimated as 0.69 and 0.73 for the total scale and PSS Function subscale, respectively, indicating that both scales were substantially reliable. PSS Pain and PSS Satisfaction subscales had moderate test-retest reliability estimates of 0.55 and 0.48 respectively for the total sample. The difference in estimates between the study by Leggin and this current study could be due to differences in the time intervals between the test and retest values across the studies. Leggin had a shorter time interval of four weeks as compared to the thirteen-week interval that was selected in the current study. Boonstra and colleagues reported test-retest reliability to be between "moderate" and "good" for the VAS-Pain scale [17]. In the current study, we obtained a test-retest reliability estimate of 0.62 for VAS-Pain, a magnitude also considered "substantial"[31]. The relatively lower test-retest reliabilities in the current study could be explained due to missing data in our sample with thirty-six individuals having missing data from our sample of 162 individuals.

According to Cohen, a 'small' effect size is one in which there is a real effect which can be proved with careful study, and a 'large' effect size is an effect which can be determined even without careful study [32]. Cohen provided thresholds on the impact of an intervention, with values of 0.2 generally regarded as "small," 0.5 as "medium," and 0.8 as "large" [32]. In our study, we found excellent responsiveness of 1.51 for PCS-12, no responsiveness was found for MCS-12 (0.005). Effect size for the total Penn Shoulder Score was excellent in the current study, with a value of 2.71, while the effect size for Penn Shoulder Score as estimated by Leggin and colleagues was lower but also large (1.01). One possible reason why effect sizes for the Penn Shoulder Score were larger than those reported by Leggin and colleagues could be that the time period that Leggin et al. included was a four-week period between two successive visits by patients [13]. We included the baseline pre-operative scores and the final post-operative scores after one year to determine the responsiveness. The longer time period in our study could have contributed to higher effect sizes.

For all the scales, the effect size for individuals who suffered from retear were observed to be lower than the effect sizes for individuals who did not suffer retears. This finding is

consistent with our expectations of observing lower change in mean scores of the health-related quality of life scores for individuals who suffered retears, as they would be expected to gain lower increase in their physical functioning during their recuperation relative to individuals without retears.

In addition to standardized effect sizes, we examined standardized response means [33], which are calculated by subtracting each individual's baseline score at pre-operative baseline visit from the final visit at week 52 and dividing the mean change in score by standard deviation of change in scores (instead of the standard deviation of scores at baseline). For all scales, values for standardized response means were comparable to those of the standardized effect sizes reported here.

Effect sizes obtained from retear and no retear groups can be directly compared to determine differences, or sensitivity, between the groups. The effect size between retear and no retear groups for PCS-12 was 0.30, indicating that the effect size was between "small" and "medium" according to Cohen's criterion [32]. For MCS-12, however, the effect size between retear and no retear groups was only 0.11 and not relevant. For the PSS, the effect size between groups was estimated to be medium (or moderate) with a value of 0.54. The Pain subscale of the PSS approached a medium effect size (0.45) between groups, the Satisfaction subscale of the PSS had a medium-to-large effect size of 0.71, and the Function subscale of the PSS approached a medium effect size (0.42). The VAS-Pain had a medium effect size of 0.51 between groups.

The above trends were also reflected in the cumulative distribution function plots for PSS, PCS-12 and VAS-Pain where separation occurred for most of the distribution while the same was not observed for MCS-12. The lack of the statistical significance between the curves for the PCS-12, the total PSS, and the VAS-Pain can be explained by the imbalance in the sample size between the retear group and no retear group. Out of a total sample of 127 patients, 110 were in one group (no retear) and the rest (only 17) were in the retear group. This gross lack of balance in the sample size between groups resulted in lack of statistical significance between the cumulative distribution curves between the two groups for total PSS, VAS-Pain, and possibly the PCS-12. Although the study was not designed to provide statistical power to distinguish the groups in this respect, the direction and magnitude of the difference appears to be meaningful.

The study has at least four limitations for consideration. Missing data on the PRO measures may have affected the results. Although three-quarters (74%) of data were available for PRO measures, there is no way of knowing for sure the missing data mechanism and hence the true values with all data completed. Nonetheless, a comparison between the group with missing PRO data and the group not missing PRO data on demographic and clinical factors revealed that there was no significant difference between the groups on age, gender or race. Another limitation is that the results from the trial (given its inclusion and exclusion criteria) may not be necessarily generalized to a wider population. Also, as this clinical trial was not blinded, the assessments may have been affected by responder bias. Moreover, statements on relative efficacy between interventions cannot be made, since everyone in this study had undergone surgical repair and there was no control group.

The current study has a major strength in the evaluation of the effect of rotator cuff repair and its impact on health status of patients who underwent arthroscopic surgery. The current study included both generic measures, SF-12 and VAS-Pain instruments, and a condition-specific instrument, the PSS which has not been examined as a set of measures in previous studies.

In conclusion, while the PSS total, the function subscale of PSS, the PCS-12, MCS-12, and VAS-Pain showed moderate test-retest reliability for individuals suffering from rotator cuff tears, the Pain and Satisfaction subscales of the PSS showed only fair test-retest reliability. The PSS (total) score and its Pain and Satisfaction subscales showed high internal consistency. All the scales apart from the MCS-12 showed excellent responsiveness and discernible sensitivity according to their respective effect sizes. Effect size between retear and no retear groups were of medium or moderate effect for PSS, Pain, and Satisfaction subscales as well as for VAS-Pain. However, the effect size between retear and no retear groups was small in effect for PCS-12 and was not significant in the case of MCS-12.

ACKNOWLEDGEMENTS

This work used data from a multicenter, nonrandomized clinical trial that was conducted and sponsored by Pfizer Inc. The authors are grateful to Dr. Charles D. Petrie for his time and effort in enabling and supporting this project.

DISCLOSURE

Jyothi Menon was an employee of Pfizer, Inc during a portion of the project. Joseph C. Cappelleri, Jack Mardekian and Nicholas J.Vendetti are full-time salaried employees and stockholders of Pfizer Inc. Marko Mychaskiw was a full-time salaried employee and stockholder of Pfizer Inc. Joseph Thomas III is involved as a consultant to Pfizer, Inc. on projects unrelated to current study. None of the authors have any other conflicts or relationships that present potential conflicts of interest.

REFERENCES

1. Lovering RM, Russ DW. Fiber type composition of cadaveric human rotator cuff muscles. J Orthop Sports Phys Ther. 2008; 38(11): 674-680.
2. Lorbach O, Tompkins M. Rotator cuff: biology and current arthroscopic techniques. Knee Surg Sports Traumatol Arthrosc. 2012; 20(6): 1003-1011.
3. DeHaan AM, Axelrad TW, Kaye E, Silvestri L, Puskas B, Foster TE. Does double-row rotator cuff repair improve Functional outcome of

patients compared with single-row technique? A systematic review. Am J Sports Med. 2012; 40(5): 1176-1185.
4. Chillemi C, Petrozza V, Garro L, Sardella B, Diotallevi R, Ferrara A, Gigante A, Di Cristofano C, Castagna A, Della Rocca C. Rotator cuff re-tear or non-healing: histopathological aspects and predictive factors. Knee Surg Sports Traumatol Arthrosc. 2011; 19(9):1588-1596.
5. van der Meijden OA, Westgard P, Chandler Z, Gaskill TR, Kokmeyer D, Millett PJ. Rehabilitation after arthroscopic rotator cuff repair: current concepts review and evidence-based guidelines. Int J Sports Phys Ther. 2012; 7(2): 197-218.
6. Bishop J, Klepps S, Lo IK, Bird J, Gladstone JN, Flatow EL. Cuff integrity after arthroscopic versus open rotator cuff repair: a prospective study. J Shoulder Elbow Surg. 2006; 15(3): 290-299.
7. Tauro JC. Arthroscopic rotator cuff repair: analysis of technique and results at 2- and 3-year follow up. Arthroscopy. 1998; 14:45-51.
8. Kim JR, Cho YS, Ryu KJ, Kim JH. Clinical and radiographic outcomes after arthroscopic repair of massive rotator cuff tears using a suture bridge technique: assessment of repair integrity on magnetic resonance imaging. Am J Sports Med. 2012; 40(4):786-793.
9. Boileau P, Brassart N, Watkinson DJ, Carles M, Hatzidakis AM, Krishnan SG. Arthroscopic repair of full-thickness tears of the supraspinatus: does the tendon really heal? J Bone Joint Surg Am. 2005; 87(6): 1229-1240.
10. Chung SW, Park JS, Kim SH, Shin SH, Oh JH. Quality of life after arthroscopic rotator cuff repair: evaluation using SF-36 and an analysis of affecting clinical factors. Am J Sports Med. 2012; 40(3): 631-639.
11. Hill-Briggs F, Gary TL, Hill MN, Bone LR, Brancati FL. Health-related quality of life in urban African Americans with type 2 diabetes. J Gen Intern Med. 2002; 17(6): 412-419.
12. Kirkley A, Alvarez C, Griffin S. The development and evaluation of a disease-specific quality-of-life questionnaire for disorders of the rotator cuff: the Western Ontario Rotator Cuff Index. Clin J Sport Med. 2003; 13(2): 84-92.
13. Leggin BG, Michener LA, Shaffer MA, Brenneman SK, Iannotti JP, Williams GR Jr. The Penn shoulder score: reliability and validity. J Orthop Sports Phys Ther. 2006; 36(3):138-151.
14. Scalise JJ, Ciccone J, Iannotti JP. Clinical, radiographic, and ultrasonographic comparison of subscapularis tenotomy and lesser tuberosity osteotomy for total shoulder arthroplasty. J Bone Joint Surg Am. 2010; 92(7):1627-1634.
15. Fan ZJ, Smith CK, Silverstein BA. Responsiveness of the QuickDASH and SF-12 in workers with neck or upper extremity musculoskeletal disorders: one-year follow-up. J Occup Rehabil. 2011; 21(2): 234-243.
16. Seymour RA. The use of Pain scales in assessing the efficacy of analgesics in post-operative dental pain. Eur J Clin Pharmacol. 1982; 23(5): 441-444.
17. Boonstra AM, Schiphorst Preuper HR, Reneman MF, Posthumus JB, Stewart RE. Reliability and validity of the visual analogue scale for disability in patients with chronic musculoskeletal Pain. Int J Rehabil Res. 2008; 31(2): 165-169.
18. Pfizer Arthroscopic Surgical Outcome Study in Subjects with Rotator Cuff Tears. In: ClinicalTrials.gov [Internet]. Bethesda (MD): National Library of Medicine (US). [2013Aug10]. Available from: URL of the record NLM Identifier: NCT 00739947.
19. Iannotti JP, Deutsch A, Green A, Rudicel S, Christensen J., Marraffino S. et al. Time to failure after rotator cuff repair: a prospective imaging study. J Bone Joint Surg Am. 2013; 95(11): 965-971.
20. Ware JE, Jr, Kosinski M, Keller SD. How to Score the SF-12 Physical and Mental Health Summary Scales. 2d. Boston: The Health Institute, New England Medical Center. 1995.
21. Burckhardt CS, Kim D. Jones. Adult measures of Pain: The McGill Pain Questionnaire (MPQ), Rheumatoid Arthritis Pain Scale (RAPS), Short-Form McGill Pain Questionnaire (SF-MPQ), Verbal Descriptive Scale (VDS), Visual Analog Scale (VAS), and West Haven-Yale Multidisciplinary Pain Inventory (WHYMPI). Arthrit Care Res. 2003; S96-S104.
22. The data analysis for this paper was generated using SAS software, Version 9.3 of the SAS System for Unix. Copyright © 2011 SAS Institute Inc. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, USA.
23. StataCorp. Stata Statistical Software: Release 12. College Station, TX: StataCorp LP. 2011.
24. Nunnally, J. C. Psychometric theory (2nd ed.) New York: McGraw-Hill. 1978.
25. Hamer RM. (2002). INTRACC.SAS. 90. Cary, NC: SAS Institute:http://support.sas.com/kb/25/031.html. Accessed 5 July 2013.
26. Shrout, P.E. and Fleiss, J.L. Intraclass correlations: uses in assessing rater reliability. Psychol Bull. 1979; 86:420-428.
27. Schuck P. Assessing reproducibility for interval data in health-related quality of life questionnaires: which coefficient should be used? Qual Life Res. 2004; 13(3): 571-586.
28. Cappelleri JC and Bushmakin AG. Interpretation of patient-reported outcomes. Statistical Methods in Medical Research. 2014; 23(5): 460-483.
29. McLeod, LD, Coon CD, Martin SA, Fehnel SE and Hays RD. Interpreting patient-reported outcome results: US FDA guidance and emerging methods. Expert Rev Pharmacoecon Outcomes. 2011; 11:163-169.
30. Pappas PA, DePuy V. An overview of non-parametric tests in SAS: when, why, and how. Paper TU04. Duke Clinical Research Institute, Durham. 2004; 1–5.
31. Landis, J.R., and G.G. Koch. The measurement of observer agreement for categorical data. Biometrics. 1977; 33:159-174.
32. Cohen, J. A power primer. Psychol Bull. 1992; 112(1):155–159.
33. Husted JA, Cook RJ, Farewell VT, Gladman DD. Methods for assessing responsiveness: a critical review and recommendations. J Clin Epidemiol. 2000; 53:459 – 468.

Source of Support: Nil, Conflict of Interest: None declared